# Predicting Startup Funding Momentum with Collective Intelligence

**Tom Kehler**[*]
CrowdSmart
44 Tehama St
San Francisco, CA 941053
tom@crowdsmart.io

**Sonali Sinha**
CrowdSmart
44 Tehama St
San Francisco, CA 941053
sonali@crowdsmart.io

**Fengchen Liu**
CrowdSmart
44 Tehama St
San Francisco, CA 941053
fengchen@crowdsmart.io

**Matt Olfat**
CrowdSmart
44 Tehama St
San Francisco, CA 941053
matt@crowdsmart.io

## Abstract

In this paper we describe a set of techniques that leverage human and machine intelligence to create data and models that are both predictively accurate and support an in depth explanation facility.Our approach leverages the cognitive diversity of a evaluation teams with domain experience and investment experience. Each team, unique to each startup, participates in an interactive evaluation process where each member of the team uses available evidence to score the potential for a startup to produce a return on investment. All startups are then tracked for performance following scoring and prediction. To date the approach outlined here demonstrates high predictive accuracy.

Key to the explanatory power of our approach is collecting natural language data on reasons behind a particular score. All supporting reasons to all scores are collected and sampled to learn rank and relevance determined by group responses. Each reason is assigned a relevancy score based structured interactions with the evaluating team. A typical evaluation will consist of ~200 quantitative data items and ~10,000 words.

Thematic analysis of the supporting reasons for scores is produced by training an NLP model to map reasons into a predetermined set of themes typically used in evaluating investments. Feedback from the evaluating teams and startup founding teams allow for continuous training of explanation system.We show that an architecture of interoperable models are highly effective in achieving both accuracy and explanatory power in investment evaluations.

## 1 Introduction

Generally, collective intelligence refers to the process by which groups of diverse individuals pool their knowledge, data and skills to contribute to a prediction or solve a problem. Specifically, we are interested in utilizing a disciplined collective intelligence approach to a group's ability to predict the potential for success of a startup based on a typical set of investment materials. Our approach is based on the findings that a diverse group of well-informed diverse individuals will outperform individual

[*]Corresponding author: Dr. Tom Kehler. Email: tom@crowdsmart.io Website: https://crowdsmart.io

expert predictions [13]. A typical investment evaluation processes for early stage investing is largely based on individual relationships. Whether the investment is coming from a venture capitalist or angel investor there is a scarcity of data available other than through direct interactions with the startup. The time, effort, and quality of data collection and evaluation can have major impact on investment results. There can be as much as a 6 to 7 times better investment return performance based on the level of due diligence [20].

In this paper we explore the application of collective intelligence techniques to the data gathering and evaluation process. Each startup presents their investor packet including but not limited to an investor deck, pro forma financials, and team information. A group of informed individual (e.g. angel or institutional) investors review the materials in an asynchronous highly interactive process that runs over a couple of weeks. First the group (~25 to 30 individuals) reviews the materials and assesses the company on four broad categories:

1. How compelling is the business opportunity from a market size, product market fit, competitive perspective?
2. How would you rate the team based on the expertise required to execute the strategy?
3. How would you rate the early supporters and investors in terms of their ability to use connections and expertise to provide assistance to the company?
4. How likely are you personally to invest or strongly recommend this investment to others?

Each of these dimensions is scored on a 1 to 10 scale.

The evaluators are asked to give their reasons for their ratings. In addition, they are asked to review and rank in priority order of agreement a sample of reasons submitted by other evaluators. Evaluators do not see others sores prior to submitting their own. Also, evaluators do not know the identity of other evaluators.

The evaluators are then provided an opportunity to participate in a 60 to 90 minute Q&A session with the startup team. Evaluators are then asked to recast their votes and reasons as they learn more about the startup. All aspects of the ratings and interactions are captured for analysis. In the course of an evaluation we collect approximately 200 quantitative data points and approximately 10,000 words.

As described above, one important feature of our method is its transparency and interpretability. Recently the problem of opacity in machine learning algorithms has become a large concern, as automated decision-making processes have increasingly been given sway in real-world settings. In particular, the black-box nature of many such processes complicates maintenance and obfuscates erroneous outputs [19]. This has inspired a growing line of work among the research community on interpretable learning techniques, especially in NLP tasks [9, 10]. The earliest works in this field concerned the design of *post hoc* explainer modules, which were layered onto existing models and sought to generate explanations for the decisions that the models made [18, 6]. Among the early approaches in the most recent wave of interpretable machine learnign algorithms were those that sought to design classifiers using submethods deemed inherently interpretable [16], as well as those that sought to enforce interpretability *a priori* via specialized feature selection steps [14]. Similar in inspiration are works that have argued for the inherent interpretability of sparse models, and then sought to operationalize this in designing interpretable methods [17, 10]. These methods represented notable steps fowards in this domain, but often lacked rigorous justifications for their notions of interpretability.

Alternatively, a recent and growing strain of the literature has instead focused on attention-based models, which rely on an in-built auto-encoder that only encodes critical and interpretable aspects of the data, thus narrowing the model's "attention"; such an approach has found use-cases in problems as varied as natural language processing [2], computer vision [24, 21], and image captioning [8, 22, 23], and has even been merged with the aforementioned pre-processing techniques to improve flexibility [15]. In effect, these approaches take advantage of the physical nature of the input data as they seek to restrict "attention". Critically, rigorous mathematical definitions of interpretability have been also proposed and argued for in recent work [1]. While presenting significant gains in empirical results and theoretical justifications over previous work, these new approaches are often highly data intensive and possibly inflexible. The data associated with the problem that we are faced with does not easily fit into many of the frameworks described, and so requires a different approach.

# 2 Methodology

## 2.1 Collective Bayesian Learning

Each evaluator is modelled as a Bayesian learner. That is, we assume they will update their scores based on evidence. The system is transactional and supports a time series of assessments during the course of evaluating a particular startup. Evaluators are encouraged to re-score companies based on new information that surfaces from the interactive evaluation process.

## 2.2 Relevance Scoring: Learning Points of Agreement From the Evaluators

A key factor in an investment evaluation is to detect signals that imply the group is leaning towards an "invest" decision or a "pass" decision. To do that we use an algorithm that learns the rank and relevance of reasons from the evaluators using a Markov sampling process. In general given a group of $N$ collaborators we define a process that creates a rank ordered list of reasons or ideas using Markov sampling [12].

In developing the algorithm we have the following objectives:

1. The process must give high ranking to the ideas that are viewed as most important to the crowd of participants.

2. The process must be repeatable when you run the process with the same problem or question with the similar set of participants.

3. The process must conserve sample. Specifically the goal is to achieve convergence with the minimum set of rating events.

The details of the specific sampling and scoring process along with simulations to test conversion rates are covered elsewhere [7]. The conclusion is that the approach is an accurate way to both collect and rank order ideas in response to a stimulus.

## 2.3 The Evaluation Data Set

The context of all data collection is a startup $S$ and a collection of evaluation assets. For each $S$ there is an evaluation team $E = \{e_1, e_2, ..., e_n\}$ with each evaluator having a set of quantitative scores $s = \{s_1, s_2, ..., s_k\}$ and a set of reasons $r = \{r_1, r_2, ..., r_j\}$. Each reason includes the following attributes: text, author, rank, score, supporter list, associated quant score etc. The attribute score: $r_{score}$ , is best viewed as the reasons "batting average". When sampled and viewed how many times did it "win" in its priority position over other ideas in the sample. *Reasons are interpreted as the drivers behind the quantitative ratings.*

The universe of reasons for a score spans all quantitative features measured. Specifically, an evaluator may decide to offer a reason for a high score in market opportunity that includes text discussing the team. The universe of reasons is segmented into a set of themes, $t = \{t_1, t_2, ..., t_l\}$.

Before going further we show an example from the platform. Each measured feature with have a distribution based on the scoring pattern. For example: In response to "How likely are you to invest in $S$?" we have the distribution shown in Figure **??** .

In addition to the data distribution, the dashboard has the reasons presented in order of their priority score, $r_{score}$ as well as its ranking in the universe of reasons. Reasons can be replied to by the startup permitting a dialog around issues. Evaluators can respond to the reasons and be influenced by the answers provided by the startup team.

Themes are collections of $\{r\}$ with $Max\{r_{score}\}$ determining $t_{score}$.The theme analysis of this particular startup shows themes touching multiple features within the evaluation. In the example shown below, we show that the theme "Technical and Research Team" spans the market opportunity, team assessment and likely to invest. $t_{score}$= 75 and there are 12 reasons assigned to this theme.

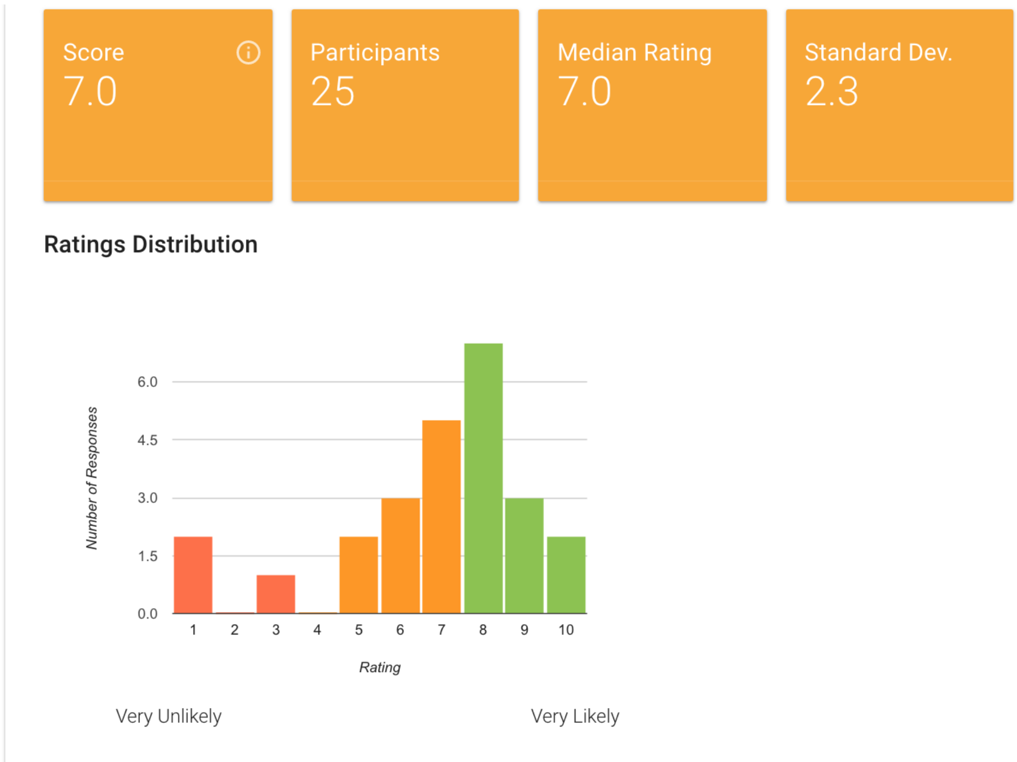


Figure 1: Feature Score Distribution

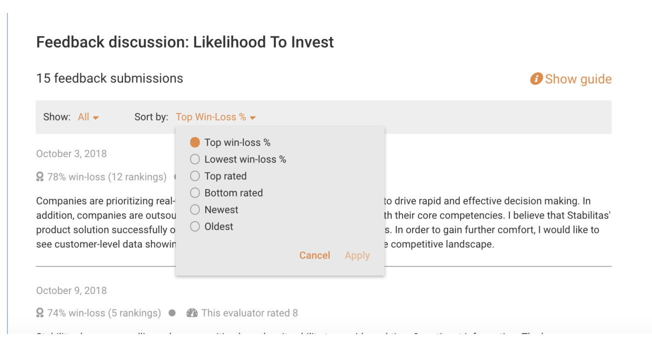


Figure 2: Reasons driving scores

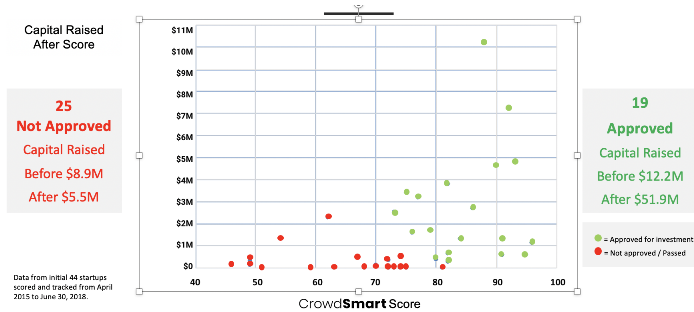


Figure 3: Funding Momentum Results

# 3 Analysis of the Data

## 3.1 Analysis of Numeric Scores

Each startup has a data set generated from evaluation sessions that consist of ~200 quantitative data items and ~10,000 words. Quantitative scores are evaluated via a logistic regression model that is trained on investment outcomes. Evaluators are potential investors. A significant portion of the evaluators actually make investments. Their intent in the evaluation process is to determine if this is a company where they will make an investment. Specifically, they are evaluating with the expectation of achieving some form of ROI. Thus we interpret the probability of assignment to the "invest" class as a probability of an ROI (pROI). Each startup receives a pROI. If that number is solidly in the invest zone (e.g. pROI > 80%), an investment is made.

## 3.2 Analysis of the scored text data

In addition to the numerical scores, substantial insights come in the language portion of the answers which we label as reasons (for the score). As discussed previously in addition to each reason being associated with a quantitative score, the adaptive learning algorithm assigns a relevancy score to each reason. Reasons are assigned a theme by the NLP classifier with each theme having a relevancy score. The theme relevancy score is equal to the reason with the maximum relevancy score in the theme class. The theme names are taken from a knowledge model based on interviews and research on what topics investors consider in making an investment decision; this theme list will grow over time. Some examples of themes are "vision of the company", "performance of prototype", "technological innovation" and others. Because we are collecting both the quant score and qualitative feedback, there is context for this data. If we plot a distribution of quant scores within each theme, then interesting patterns start to emerge. Some distributions are normal, others are uniform, others are skewed (left skewed and right skewed), and others are bi-modal. Thus a theme has a distribution of quantitative scores that span the features being assessed. Analysis of the underlying score data associated with the reasons allows us to label themes in terms of representing "confidence in an investment" or "reason for risk or further discussion."

## 3.3 Prediction Accuracy

All startups scored on the platform are tracked for funding results after a prediction is made. Early results are promising as indicated in results figure. The system to date has a greater than ~90%) accuracy.

# 4 Context, A Critical Foundation For Explanation

Knowledge-based systems in the first AI wave made transparent the logic of a system through explanation. Explanation provides a powerful tool for guiding a system to semantically meaningful results. Explicit representation of domain knowledge provides the foundation for generating explanations. The use of domain knowledge to frame reasoning in classic knowledge-based systems is covered in an early paper by Fikes and Kehler [11].

Logic, symbolic pattern recognition, and rule-based reasoning were foundational the first generation AI systems. These systems were limited by the need to explicitly program the core symbolic reasoning components. Machine learning and mathematical techniques replaced the labor intensive symbolic programming requirements with system components that learn from data. Mathematically based systems are largely black box systems. While black box systems are extremely useful, they lack in explanatory power. In the same way knowledge representation in the form of frames manages context and provides a foundation for explanation in symbolic systems, a related technique can be used to support explanation in machine learning systems. We use Bayesian belief networks as a way to dynamically create explanations utilizing the scored reasons and themes resulting from each evaluation process.

Bayesian belief networks are known by names such as causal graphs, causal networks, belief networks, recursive models, probabilistic networks or permutations of two or three of these terms [3, 5, 4]. A Bayesian belief network structure is a directed acyclic graph in which nodes represent domain variables and arcs between nodes represent probabilistic dependencies [5]. Figure 4 shows an example of a Bayesian belief network taken from the CrowdSmart prediction platform. It models the following variables: the startup's pROI (node in green); the startup's $Features = \{feature_1, feature_2, ..., feature_l\}$ (nodes in cyan); the startup's $Themes = \{theme_1, theme_2, ..., theme_m\}$ categories (nodes in light blue); and the $Reasons = \{reason_1, reason_2, ..., reason_n\}$ (nodes in grey) behind the Themes. The presence of the arc between two nodes ($feature_k$ and pROI) in the network structure implies that $feature_k$ and pROI are directly dependent, relative to the other nodes. The absence of an arc between two nodes (for example $reason_j$ and pROI) implies that $reason_j$ and pROI are dependent through $theme_i$ and the $feature_k$. The key feature of Bayesian belief networks is their explicit representation of the conditional independence among events; a Bayesian belief network represents a full joint-probability space over $n$ event variables in the network, and the joint probability can be computed as the product of only $n$ probabilities [4]. Therefore, given the CrowdSmart Bayesian belief network structure in Figure 4, the joint probability of pROI, $l$ $Features$, $m$ $Themes$, and $n$ $Reasons$ can be calculated by:

$$P(pROI, Features, Themes, Reasons) = P(pROI|feature_1, ..., feature_l) \cdot \prod_{k=1}^{l} P(feature_k|theme_{feature_k}) \cdot \prod_{i=1}^{m} P(theme_i|reason_{theme_i}) \cdot \prod_{j=1}^{n} P(reason_j)$$

In the system discussed here we use Bayesian Belief Networks as a way to frame context and support explanation. For example, at the highest level the frame for analysis and prediction is a startup $S$. $S$ is characterized by a set of assets (pitch deck, financials, team). With each startup we have four high level evaluation frames:

1. Compelling Opportunity (CO) is the context for evaluating product market fit, competition and other aspects of the market opportunity addressed by the startup.

2. Team Assessment (TA) is the context for evaluating the business, technical and developmental aspects of the startup team.

3. Investor Assessment (IA) is the context of evaluating the network support for the startup. Do they have domain and business experts as part of their early supporters and investor group.

4. Likely To Invest (LI) is the context for assessing commitment of financial resources to the startup. Specifically do you as evaluator believe enough to put skin in the game.

Each contextual frame (CO, TA, IA, LI) is a node in a Bayesian Belief Network. So in its most simple form the contextual network is shown in Figure 4.

Here we have made the representation very simple. Assessments are discrete, the state of CO is linked to the startups states (Invest or Pass) by a conditional probability table.

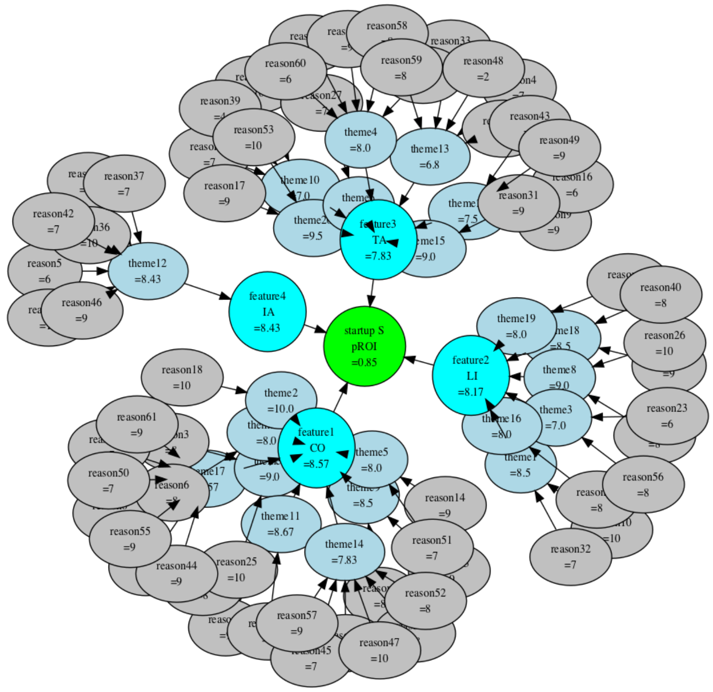


Figure 4: CrowdSmart Bayesian Belief Network Structure

## 5 Explanation Models in Financial Predictions

### 5.1 Explanations as a foundation for trust

In order to facilitate bayesian learning by the evaluators we utilize the relevance scoring of reasons and themes to surface preliminary results as an evaluation proceeds. Thus an evaluator can see a summary of key drivers of scores before the evaluation is completed and a pROI is generated. These preliminary results are shared with the evaluating team and startup team in a recorded interactive session. By this mechanism we are able to check system explanations and analysis with human evaluator and startup response to system explanation summaries.

The explanation system supports a mutual learning environment in which startups and evaluators can learn from each other and work together to create an outcome (e.g. invest or pass based on the pROI) which has transparent logic. That is the explanation model parallels the numeric outcome of a pROI.

An example result is shown in the attached Figure 6 and Figure 5. The startup has a pROI and an associated explanation.

### 5.2 Analysis of Drivers

A theme with a normal distribution around a favorable score (e.g.) 8 is interpreted as a driver to a score leading to investment. Specifically we could read the result that these reasons are driving higher scores on market or team assessment. Sentiment analysis provides a parallel and perhaps overlapping signal to interpret drivers. If a theme has a high sentiment analysis score and the mean theme score is high as well, we have further evidence supporting interpretation of drivers to an invest or pass decision. A next level of analysis involves looking at the associated theme distribution characteristics and creating inferences on driving patterns.

The results show that single properties of the distribution, like mean, mode, variance, or even sentiment analysis, don't fully capture the concerns raised in the feedback. However, if every distribution is seen as a mixture of two separate normal distributions, one part being a risk component and one part being a reward component, then concerns that experts express in the feedback show in

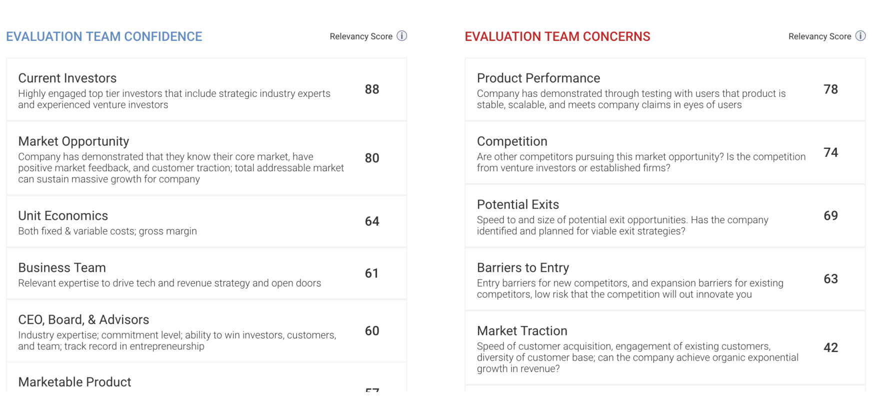


Figure 5: Themes analysis

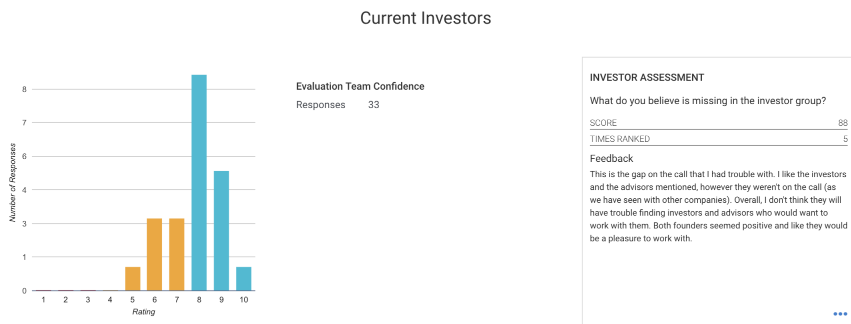


Figure 6: Current investors

the overall pattern of the distributions. Thus, if mean - mode is above a threshold, then that correlates well with the language and sentiment of the feedback.

While this is a small dataset it indicates consistency and establishes a foundation for exploring drivers of decision convergence. Now that we can zero in on what the investor concerns are, the startup and investors can have a discussion around themes. The startup then presents new information to answer concerns, and evaluators can reevaluate.

## 6 Summary and Future Work

The model generates predictions based on collective intelligence accompanied by explanations that are intended to build trust in the predictions. We propose that investment decision can be made based on machine learning model but supported by trust building explanations. Interoperable, parallel knowledge models and numeric models are required to accomplish this goal. We have shown that to date the accuracy of predictions tracks nicely to outcomes. We have also shown that the explanation facility elucidates the "why" behind the predictions.

We continue to explore parallel interoperable models as data sets of scored companies. We also continue to track companies as they develop to improve and train our models. In particular, we are conducting experiments with the knowledge model to infer predictive values for pROI from natural language analysis alone. Specifically, initial results are promising that our approach serves as a new type of "sentiment analysis" that can infer pROI probability estimates by running discussions about a startup through the BBN-based knowledge model.

## References


[1] David Alvarez-Melis and Tommi S Jaakkola. Towards robust interpretability with self-explaining neural networks. *arXiv preprint arXiv:1806.07538*, 2018.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[3] Remco Ronaldus Bouckaert. *Bayesian belief networks: from construction to inference*. PhD thesis, 1995.

[4] Gregory F Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial intelligence*, 42(2-3):393–405, 1990.

[5] Gregory F Cooper and Edward Herskovits. A bayesian method for constructing bayesian belief networks from databases. In *Uncertainty Proceedings 1991*, pages 86–94. Elsevier, 1991.

[6] Mark Craven and Jude W Shavlik. Extracting tree-structured representations of trained networks. In *Advances in neural information processing systems*, pages 24–30, 1996.

[7] CrowdSmart. Unpublished crowdsmart white paper. 2018.

[8] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, et al. From captions to visual concepts and back. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1473–1482, 2015.

[9] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014.

[10] Manaal Faruqui, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah Smith. Sparse overcomplete word vector representations. *arXiv preprint arXiv:1506.02004*, 2015.

[11] Richard Fikes and Tom Kehler. The role of frame-based representation in reasoning. *Communications of the ACM*, 28(9):904–920, 1985.

[12] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

[13] Lu Hong and Scott E Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.

[14] Been Kim, Julie A Shah, and Finale Doshi-Velez. Mind the gap: A generative approach to interpretable feature selection and extraction. In *Advances in Neural Information Processing Systems*, pages 2260–2268, 2015.

[15] Tao Lei, Regina Barzilay, and Tommi Jaakkola. Rationalizing neural predictions. *arXiv preprint arXiv:1606.04155*, 2016.

[16] Benjamin Letham, Cynthia Rudin, Tyler H McCormick, David Madigan, et al. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, 9(3):1350–1371, 2015.

[17] Brian Murphy, Partha Talukdar, and Tom Mitchell. Learning effective and interpretable semantic models using non-negative sparse embedding. *Proceedings of COLING 2012*, pages 1933–1950, 2012.

[18] Sebastian Thrun. Extracting rules from artificial neural networks with distributed representations. In *Advances in neural information processing systems*, pages 505–512, 1995.

[19] Alfredo Vellido, José David Martín-Guerrero, and Paulo JG Lisboa. Making machine learning models interpretable. In *ESANN*, volume 12, pages 163–172. Citeseer, 2012.

[20] Robert Wiltbank and Warren Boeker. Returns to angel investors in groups. 2007.

[21] Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, and Zheng Zhang. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 842–850, 2015.

[22] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057, 2015.

[23] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola. Stacked attention networks for image question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 21–29, 2016.

[24] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.